

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-045267

(43)Date of publication of application : 16.02.1999

(51)Int.Cl.

G06F 17/30

(21)Application number : 09-201984

(71)Applicant : JUST SYST CORP

(22)Date of filing : 28.07.1997

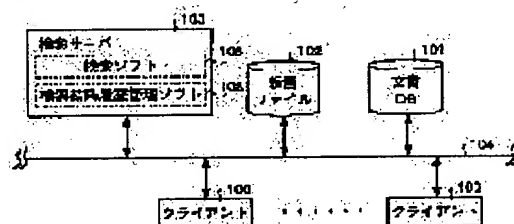
(72)Inventor : UKIKAWA KAZUNOBU
UKIKAWA HATSUKO
TAKATO ATSUSHI
ARAI YOSHIYA

(54) DOCUMENT RETRIEVAL DEVICE AND COMPUTER READABLE RECORDING MEDIUM RECORDED WITH PROGRAM FOR FUNCTIONING COMPUTER AS THE DEVICE

(57)Abstract:

PROBLEM TO BE SOLVED: To contribute to information analysis and the creation of a new value.

SOLUTION: A retrieval server 103 is provided with a retrieval software 105 for inputting retrieval conditions from a client 100, retrieving a pertinent document, extracting a noun phrase from the document of a retrieved result, imparting a score corresponding to statistic information such as an appearing frequency and distribution, etc., in the document of the retrieved result and the document of a retrieval object and extracting the noun phrase of the score pertinent to extraction conditions set beforehand as the relating word of the retrieval conditions and a retrieved result history management software 106 for storing and managing at least the retrieval conditions, the retrieved result by a retrieval means, the relating word extracted in a relating word extraction means and a date on which the retrieval is performed as retrieved result history. In this case, the client 100 reads the retrieved result history managed by the retrieved result history management software, displays it on a screen and performs the retrieval and analyzes information by utilizing the retrieved result history.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-45267

(43) 公開日 平成11年(1999) 2月16日

(51) Int.Cl.⁴
G 0 6 F 17/30

識別記号

F I
G 0 6 F 15/403 3 5 0 C
15/40 3 7 0 A
15/403 3 4 0 B

審査請求 未請求 請求項の数 2 O L (全 10 頁)

(21) 出願番号 特願平9-201984

(22) 出願日 平成9年(1997) 7月28日

(71) 出願人 390024350

株式会社ジャストシステム
徳島県徳島市沖浜東3-46

(72) 発明者 浮川 和宣

徳島市沖浜東3丁目46番地 株式会社ジャ
ストシステム内

(72) 発明者 浮川 初子

徳島市沖浜東3丁目46番地 株式会社ジャ
ストシステム内

(72) 発明者 高藤 淳

徳島市沖浜東3丁目46番地 株式会社ジャ
ストシステム内

(74) 代理人 弁理士 酒井 昭徳

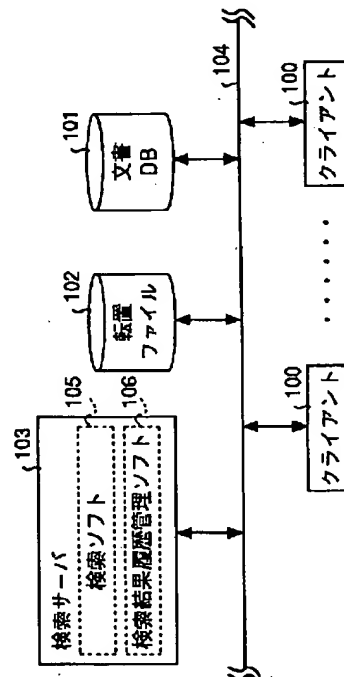
最終頁に続く

(54) 【発明の名称】 文書検索装置およびその装置としてコンピュータを機能させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体

(57) 【要約】

【課題】 情報分析や新たな価値の創出に寄与することができるようにすること。

【解決手段】 検索サーバ103は、クライアント100から検索条件を入力して該当する文書を検索し、検索結果の文書から名詞句を抽出し、抽出した名詞句に対して、検索結果の文書および検索対象の文書における出現頻度および分布等の統計情報に応じたスコアを付与し、予め設定された抽出条件に該当するスコアの名詞句を検索条件の関連語として抽出する検索ソフト105と、少なくとも、検索条件、検索手段による検索結果、関連語抽出手段で抽出した関連語および検索を行った日付を検索結果履歴として記憶して管理する検索結果履歴管理ソフト106と、を備えており、クライアント100は、検索結果履歴管理ソフト106で管理された検索結果履歴を読み出して画面表示し、この検索結果履歴を利用して検索を行ったり、情報の分析を行ったりすることができる。



【 特許請求の範囲】

【 請求項1 】 検索条件に基づいて、検索対象の文書から該当する文書を検索する文書検索装置において、前記検索条件を入力し、入力した検索条件に基づいて、該当する文書を検索する検索手段と、

前記検索手段で検索した結果の文書から名詞句を抽出し、抽出した名詞句に対して、前記検索手段で検索した結果の文書および検索対象の文書における出現頻度および分布等の統計情報に応じたスコアを付与するスコア付与手段と、

前記スコア付与手段で付与したスコアに基づいて、予め設定された抽出条件に該当するスコアの名詞句を前記検索条件の関連語として抽出する関連語抽出手段と、

前記入力した検索条件、検索手段による検索結果、関連語抽出手段で抽出した関連語および検索を行った日付を少なくとも検索結果履歴として記憶する履歴記憶手段と、

前記履歴記憶手段に記憶された検索結果履歴を読み出して画面表示する表示手段と、を備えたことを特徴とする文書検索装置。

【 請求項2 】 前記請求項1 に記載の文書検索装置の各手段としてコンピュータを機能させるためのプログラムを記録したことを特徴とするコンピュータ読み取り可能な記録媒体。

【 発明の詳細な説明】

【 0 0 0 1 】

【 発明の属する技術分野】 本発明は、検索を行った後、少なくとも、検索条件、検索結果、関連語および検索を行った日付を検索結果履歴として記憶しておき、記憶した検索結果履歴を後の検索処理に利用することができるようにして、情報分析や新たな価値の創出に寄与することができるようにした文書検索装置およびその装置としてコンピュータを機能させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体に関する。

【 0 0 0 2 】

【 従来技術】 複数の文書を格納した文書DB（データベース）から特定の文書を検索する文書検索装置は、一般に、検索式や検索文等の検索条件を入力し、入力した検索条件に該当する文書を文書DBから検索するものである。

【 0 0 0 3 】 ところで、文書検索装置で一度検索を行った後、同一または類似の検索条件で再度検索を行うには再び検索条件を入力しなければならなかったことから、検索を行う毎に検索条件や検索結果を保存できるようにし、次回検索を行う際に、保存しておいた検索条件や検索結果を利用することができるようにした文書検索装置が提案されている。

【 0 0 0 4 】

【 発明が解決しようとする課題】 しかしながら、上記従来の技術においては、単に検索条件や検索結果を保存す

るだけであるため、検索結果を元にして情報の分析を行う場合等には使い勝手が悪いという問題点があった。特に、検索結果を定点観測的な比較分析に利用する場合には、ユーザが独自に分析のためのデータの整理および管理を行う必要があった。

【 0 0 0 5 】 本発明は上記に鑑みてなされたものであって、検索を行った後、少なくとも、検索条件、検索結果、関連語および検索を行った日付を検索結果履歴として記憶しておき、記憶した検索結果履歴を後の検索処理に利用することができるようにして、情報分析や新たな価値の創出に寄与することができるようにすることを目的とする。

【 0 0 0 6 】

【 課題を解決するための手段】 上記目的を達成するため、請求項1 の文書検索装置は、検索条件に基づいて、検索対象の文書から該当する文書を検索する文書検索装置において、前記検索条件を入力し、入力した検索条件に基づいて、該当する文書を検索する検索手段と、前記検索手段で検索した結果の文書から名詞句を抽出し、抽出した名詞句に対して、前記検索手段で検索した結果の文書および検索対象の文書における出現頻度および分布等の統計情報に応じたスコアを付与するスコア付与手段と、前記スコア付与手段で付与したスコアに基づいて、予め設定された抽出条件に該当するスコアの名詞句を前記検索条件の関連語として抽出する関連語抽出手段と、前記入力した検索条件、検索手段による検索結果、関連語抽出手段で抽出した関連語および検索を行った日付を少なくとも検索結果履歴として記憶する履歴記憶手段と、前記履歴記憶手段に記憶された検索結果履歴を読み出して画面表示する表示手段と、を備えたものである。

【 0 0 0 7 】 また、請求項2 のコンピュータ読み取り可能な記録媒体は、前記請求項1 に記載の文書検索装置の各手段としてコンピュータを機能させるためのプログラムを記録したものである。

【 0 0 0 8 】

【 発明の実施の形態】 以下、本発明の文書検索装置およびその装置としてコンピュータを機能させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体の実施の形態について、添付の図面を参照しつつ詳細に説明する。

【 0 0 0 9 】 図1 は、本実施の形態の文書検索装置のシステム構成図である。図1 に示す文書検索装置は、所望の文書を検索するための検索条件を出力する複数のクライアント1 0 0 と、文書DB（データベース）1 0 1 の文書を用いて転置ファイル1 0 2 を生成すると共に、ベクトル空間法を用いた検索を行う検索エンジン（例えば、CLARITECH社のCLARIT等）を用いて、クライアント1 0 0 から検索条件を入力し、入力した検索条件に基づいて、転置ファイル1 0 2 から該当する文書を検索し、さらに、検索結果の文書から検索条件

の関連語を抽出する検索ソフト105および検索ソフト105による検索結果の履歴を管理する検索結果履歴管理ソフト106を備えた検索サーバ103と、上記クライアント100や検索サーバ103等を接続するネットワーク104と、から構成されている。

【0010】図1において、文書DB101は、クライアント100等で作成された複数の文書を格納したものであり、格納される文書は、ワープロ文書や、SGML、HTML等の構造化文書等、いかなる種類の文書であつても良い。本実施の形態においては、文書DB101に格納された文書を検索対象とするが、検索対象を文書DB101中の文書に限定するものではない。

【0011】転置ファイル102は、文書DB101中の複数の文書と、これら複数の文書それぞれから後述する方法で抽出した複数の索引語との関係を規定することにより、ある索引語が各文書それぞれにおいてどの程度重要な語彙であるかをベクター表現を用いて示したものであつて、この索引語を用いて該当する文書を検索することができるようにしたものである。

【0012】具体的には、1つの文書を予め定めた複数のセンテンスからなるサブドキュメント単位に区切り、サブドキュメントから上記索引語となる名詞句を抽出して、抽出した名詞句それぞれについて、サブドキュメント中の出現頻度、文書DB101全体における分布等の統計情報を求め、求めた名詞句毎の統計情報を用いて各サブドキュメントをベクター表現に変換する。そして、変換したサブドキュメントのベクター表現に基づいて、文書のベクター表現を生成する。転置ファイル102は、このようにしてベクター表現された文書DB101中の文書を格納するものである。

【0013】なお、各索引語には、対応する文書中の重要度に応じた重み付けを行うことができる。また、文書のベクター表現については、実際の検索を行う際に、サブドキュメントのベクター表現に基づいて生成することにしても良い。

【0014】クライアント100および検索サーバ103は、パーソナルコンピュータやワークステーション等によって構成される。図2は、検索サーバ103における検索ソフト105の処理を示す概略ブロック図である。検索サーバソフト105は、文書DB101中の文書を転置ファイル102に登録する処理と、ベクトル空間法を利用した文書の検索処理と、検索結果の文書から検索条件の関連語を抽出する処理を行うものである。

【0015】検索ソフト105において、文書DB101中の文書を転置ファイル102に登録する処理は、自然言語処理モジュール200と、データベース・ビルド・コンポーネント205とによって行われる。

【0016】具体的に、自然言語処理モジュール200は、文書DB101から文書を入力し、入力した文書について、フォーマットの認識処理や、品詞情報等を格納

した辞書201および各単語の係り受け等を解析するための文法辞書202を用いて形態素解析、構文解析、名詞句抽出等の解析処理を行い、上述したサブドキュメント毎の名詞句リストを含むドキュメント・セット204を生成する。

【0017】データベース・ビルド・コンポーネント205は、自然言語処理モジュール200で生成したドキュメント・セット204を入力し、入力したドキュメント・セット204中の各サブドキュメントをベクター表現に変換すると共に、サブドキュメントのベクター表現に基づいて、文書のベクター表現を生成して転置ファイル102に登録する。

【0018】また、検索ソフト105において、文書の検索処理は、自然言語処理モジュール200と、クエリー・ビルド・コンポーネント207と、検索エンジン209とによって行われる。

【0019】具体的に、自然言語処理モジュール200は、クライアント100から検索条件206を入力し、品詞情報等を格納した辞書201および各単語の係り受け等を解析するための文法辞書202を用いて形態素解析、構文解析、名詞句抽出等の解析処理を行い、検索条件206の名詞句のリストを含むドキュメント・セット204を生成する。

【0020】クエリー・ビルド・コンポーネント207は、ドキュメント・セット204を入力し、検索条件206を構成する各名詞句について、検索条件206（ドキュメント・セット204）中の出現頻度、文書DB101（転置ファイル102）全体における分布等の統計情報を求め、求めた統計情報を用いて検索条件206をベクター表現に変換したクエリー・ドキュメント208を生成する。

【0021】検索エンジン209は、クエリー・ビルド・コンポーネント207で生成したクエリー・ドキュメント208を入力し、転置ファイル102中の各文書のベクター表現とクエリー・ドキュメント208（検索条件206のベクター表現）とを比較して、クエリー・ドキュメント208との類似度に応じたスコアを各文書に付与し、所定の閾値を超えるスコアが付与された文書リスト210を出力する。

【0022】さらに、検索ソフト105において、関連語の抽出処理は、自然言語処理モジュール200と、関連語抽出エンジン211とによって行われる。

【0023】具体的に、自然言語処理モジュール200は、クライアント100で文書リスト210から文書を選択した結果を検索結果として入力し、文書DB101から検索結果に該当する文書を入力する。そして、入力した全ての文書について、フォーマットの認識処理や、品詞情報等を格納した辞書201および各単語の係り受け等を解析するための文法辞書202を用いて形態素解析、構文解析、名詞句抽出等の解析処理を行い、上述し

たサブドキュメント毎の名詞句リストを含むドキュメント・セット204を生成する。

【0024】関連語抽出エンジン211は、自然言語処理モジュール200で生成したドキュメント・セット204を入力し、入力したドキュメント・セット204中の各名詞句それぞれについて、各文書(ドキュメント・セット204)中の出現頻度や文書DB101(転置ファイル102)中の分布等の統計データを演算し、演算した統計データに基づいて、各名詞句にスコアを付与する。そして、予め設定した閾値を超えるスコアの名詞句を検索条件206の関連語句として抽出し、抽出した結果を関連語リスト212として出力する。

【0025】また、検索サーバ103において、検索結果履歴ソフト106は、検索を行った後、検索条件、検索対象、検索結果、関連語、関連文書、日付、検索者、コメント等を検索結果履歴として記憶しておき、クライアント100の要求に応じて、クライアント100のユーザが参照することができるように管理するものである。

【0026】なお、図1においては、文書DB101および転置ファイル102をネットワーク104に単独に接続した構成を示したが、これらを検索サーバ103に直接接続する構成としても良い。また、図1においては、本実施の形態の文書検索装置をネットワーク104を介したシステムで構成するように示したが、クライアント100と検索サーバ103の処理を1つのコンピュータで行うようにすることもできる。

【0027】次に、上述した構成を備えた文書検索装置の動作について、(1)転置ファイルの生成処理、(2)文書の検索処理、(3)関連語の抽出処理、(4)検索結果履歴の利用の順で詳細に説明する。

【0028】(1)転置ファイルの生成処理

図3は、転置ファイルの生成処理を示すフローチャートである。検索サーバ103の検索ソフト105は、新たな文書が文書DB101に登録された場合(S301)、この文書を入力して転置ファイル102に登録するための処理を開始する(S302)。

【0029】検索ソフト105において、自然言語処理モジュール200は、ステップS302で入力した文書を解析する処理を行う(S303)。具体的には、入力した文書がワープロ文書、HTML等の構造化文書等、いかなるフォーマットの文書であるかを判定する処理を行う。その後、辞書201および文法辞書202を用いて形態素解析、係り受け等の構文解析を行い、文書を複数のサブドキュメントに区分すると共に、区分したサブドキュメントから名詞句を抽出する等の処理を行う。

【0030】そして、自然言語処理モジュール200は、ステップS303における処理の結果に基づいて、サブドキュメント毎に名詞句リストを生成し、生成した名詞句リストを含むドキュメント・セット204を生成

する(S304)。

【0031】その後、データベース・ビルド・コンポーネント205は、自然言語処理モジュール200で生成したドキュメント・セット204を入力し、文書のベクター表現を生成して転置ファイル102に登録する処理を行う(S305)。

【0032】具体的には、ドキュメント・セット204中のサブドキュメントの各名詞句を転置ファイル102の索引語として、サブドキュメント中の出現頻度、文書DB101全体における分布等の統計情報を求め、求めた名詞句毎の統計情報を用いてサブドキュメントをベクター表現に変換する。この処理をドキュメント・セット204中の全てのサブドキュメントについて行い、変換したサブドキュメントのベクター表現に基づいて、文書のベクター表現を生成して転置ファイル102に登録する。その結果、文書DB101に新たに登録された文書がベクター表現に変換されて転置ファイル102に登録されることになり、転置ファイル102は常に最新の状態に保たれる。

【0033】(2)文書の検索処理

次に、上述したようにして生成した転置ファイル102に基づいて、文書を検索する処理について説明する。図4は、文書の検索処理を示すフローチャートである。

【0034】検索サーバ103の検索ソフト105は、クライアント100から検索条件206を入力すると(S401)、自然言語処理モジュール200において、検索条件206の解析処理を行う(S402)。この検索条件206は、自然言語で記述された検索文であり、文書検索装置の構成により、検索式、キーワードの集合等に変更することも可能である。自然言語処理モジュール200は、具体的に、検索条件206について、辞書201および文法辞書202を用いて形態素解析、係り受け等の構文解析処理を行い、検索条件206から名詞句を抽出する処理を行う。

【0035】そして、自然言語処理モジュール200は、ステップS402で解析処理を行うことによって抽出した名詞句からなるドキュメント・セット204を生成する(S403)。

【0036】続いて、クエリー・ビルド・コンポーネント207は、自然言語処理モジュール200からドキュメント・セット204を入力し、入力したドキュメント・セット204を構成する各名詞句について、検索条件206(ドキュメント・セット204)中の出現頻度、文書DB101(転置ファイル102)全体における分布等の統計情報を求め、求めた統計情報を用いてドキュメント・セット204をベクター表現に変換したクエリー・ドキュメント208を生成する(S404)。

【0037】検索エンジン209は、クエリー・ビルド・コンポーネント207で生成したクエリー・ドキュメント208を入力し、転置ファイル102中の各文書の

10

20

30

40

50

ベクター表現とクエリー・ドキュメント208(検索条件206のベクター表現)とを比較して、クエリー・ドキュメント208との類似度に応じたスコアを各文書に付与する(S405)。すなわち、ベクトル空間法を用いた検索処理が行われる。

【0038】なお、類似度に応じたスコアは、各文書とクエリー・ドキュメント208との類似度を余弦距離に基づいて表現したものであり、スコアが大きい文書がよりクエリー・ドキュメント208と類似していることを表している。

【0039】そして、検索エンジン209は、予め設定されたスコアの閾値に基づいて、閾値を超えるスコアが付与された文書を選択し、選択した文書に基づいて、文書リスト210を生成してクライアント100に出力する(S406)。

【0040】クライアント100は、検索サーバ103から文書リスト210を入力し、入力した文書リスト210に基づいて、検索された文書一覧を画面表示する(S407)。クライアント100のユーザは、画面表示された文書一覧から所望の文書を選択することにより、文書DB101中の文書を画面表示させることができる。

【0041】なお、クライアント100においては、上位のランキングの文書から順に一覧表示される。したがって、検索条件206に最も類似する文書から順に表示されることになり、ユーザが文書を選択する際の基準を提供することができる。

【0042】(3) 関連語の抽出処理

続いて、上述した検索処理において検索した文書から検索条件206の関連語を抽出する処理について説明する。図5は、関連語の抽出処理を示すフローチャートである。

【0043】クライアント100のユーザは、画面表示された文書一覧から検索結果としてふさわしい文書を選択し、選択した文書を検索結果として検索サーバ103に出力する。検索サーバ103の検索ソフト105は、クライアント100から検索結果を入力すると(S501)、検索結果に該当する文書を文書DB101から入力する(S502)。

【0044】なお、後に詳細に説明するが、クライアント100から入力した検索結果は、検索結果履歴管理ソフトにおいて検索結果履歴として記憶され、管理されることになる。

【0045】文書DB101から文書を入力すると、自然言語処理モジュール200は、入力した全ての文書について、フォーマットの認識処理や、品詞情報等を格納した辞書201および各単語の係り受け等を解析するための文法辞書202を用いて形態素解析、構文解析、名詞句抽出等の解析処理を行う(S503)。

【0046】その後、ステップS503における解析処

理の結果に基づいて、サブドキュメント毎の名詞句リストを含むドキュメント・セット204を生成する(S504)。

【0047】関連語抽出エンジン211は、自然言語処理モジュール200で生成したドキュメント・セット204を入力し、入力したドキュメント・セット204中の各名詞句それぞれについて、各文書(ドキュメント・セット204)中の出現頻度や文書DB101(転置ファイル102)中の分布等の統計データを演算する(S505)。

【0048】ステップS505で統計データを演算した後、関連語抽出エンジン211は、求めた統計データに基づいて、各名詞句に対してスコア付けを行う(S506)。このスコアは、文書200における各名詞句の重要性および検索条件206中の名詞句に対する関連性を表すもので、スコアが大きいもの程、重要性および関連性が高いことを表している。

【0049】関連語抽出エンジン211は、ステップS506で行ったスコア付けの結果に基づいて、予め設定された閾値を超えるスコアの名詞句を検索条件206中の名詞句に関連する関連語として抽出する(S507)。なお、ここでは、関連語を抽出する条件として閾値を用いることにしたが、閾値に代えて、例えば、上位5番までのスコアの名詞句を抽出することにしても良い。

【0050】その後、関連語抽出エンジン211は、ステップS507で抽出した関連語からなる関連語リスト212を生成して、クライアント100に出力する(S508)。

【0051】クライアント100は、検索サーバ103から関連語リスト212を入力し、抽出された関連語を一覧表示する。クライアント100のユーザは、一覧表示された関連語から検索結果履歴として記憶させる関連語を選択して、検索サーバ103に出力する。

【0052】(4) 検索結果履歴の利用

さらに、上述したようにして文書を検索した結果および検索結果の文書から抽出した関連語等を検索結果履歴として記憶し、記憶した検索結果履歴をいかにして利用するかについて説明する。

【0053】検索サーバ103の検索結果履歴管理ソフト106は、クライアント100から入力した検索結果(図5のステップS501参照)および関連語の選択結果、さらには、検索を行った日付やユーザ名等のユーザ情報等を用いて、検索結果履歴を作成して記憶する。

【0054】図6は、検索結果履歴の一例を示す説明図である。図6に示す検索結果履歴600には、検索に用いた検索条件、検索対象、関連語、検索結果、関連文書、日付、検索者等の欄が設けられている。また、この検索結果履歴600には、どのような意図で検索を行ったかを記録しておくためのコメント欄が設けられてお

り、クライアント100からコメントを書き込むことができるようになってい

【0055】上述したように、文書DB101に新たな文書が登録される毎に、転置ファイル102は常に更新されることになる。したがって、同一の検索条件206を用いて検索を行った場合であっても、最新の検索結果と一年前の検索結果は異なることがあり、この間の定点観測的な比較分析により、予測やリスク管理等の状況を見出せることがある。

【0056】例えば、同一トピックに関する新聞記事の経時変化(検索結果)とその時々

【0057】このように、検索結果履歴管理ソフト106で作成し、記憶した検索結果履歴600は、クライアント100から必要に応じて読み出すことができ、クライアント100のユーザは、この検索結果履歴600を利用して、検索処理に活用することができると共に、最新の検索結果と過去の検索結果履歴とを利用して、情報の分析を行うことができる。

【0058】なお、図6に示してあるように、検索結果履歴600には、同一の検索条件206を用いて定期的に検索を行うことができるように、自動検索条件欄が設けられている。ここに検索を自動的に行う時期等の条件を書き込んでおくことにより、検索ソフト105に定期的にかつ自動的に検索を行わせることができる。

【0059】以上のように、本実施の形態の文書検索装置によれば、検索を行った後、少なくとも、検索条件、検索結果、関連語および検索を行った日付を検索結果履歴として記憶しておき、記憶した検索結果履歴を利用して検索処理を行うことができるようにすることにより、情報分析や新たな価値の創出に寄与することができる。

【0060】なお、検索結果として一覧表示された文書において、ユーザが検索結果としてふさわしいと思う文書やふさわしくないと思う文書については、その結果を検索サーバ103にフィードバックすることができる。すなわち、ユーザは、検索結果としてふさわしいと思う文書に対して、正の重み、例えば「+」を指定することができ、検索結果としてふさわしくないと思う文書に対して負の重み、例えば「-」を指定することができる。その結果、入力した重みが正の指定である場合には、転置ファイル102中の該当する文書の重みが強化され、入力した重みが負の指定である場合には、文書の重みが弱められる。

【0061】また、本実施の形態においては、ベクトル

空間法による検索を例にとりて説明したが、ブーリアン検索により検索処理を行うことにしても良い。

【0062】さらに、本実施の形態で説明した文書検索装置は、予め用意されたプログラムをコンピュータやワークステーションで実行することによって実現される。このプログラムは、ハードディスク、フロッピーディスク、CD-ROM、MO、DVD等のコンピュータで読み取り可能な記録媒体に記録され、コンピュータによって記録媒体から読み出されることによって実行される。また、このプログラムは、上記記録媒体を介して、またはネットワークを介して配布することができる。

【0063】

【発明の効果】以上説明したように、本発明の文書検索装置(請求項1)によれば、検索条件を入力し、入力した検索条件に基づいて、該当する文書を検索する検索手段と、検索手段で検索した結果の文書から名詞句を抽出し、抽出した名詞句に対して、検索手段で検索した結果の文書および検索対象の文書における出現頻度および分布等の統計情報に応じたスコアを付与するスコア付与手段と、スコア付与手段で付与したスコアに基づいて、予め設定された抽出条件に該当するスコアの名詞句を検索条件の関連語として抽出する関連語抽出手段と、入力した検索条件、検索手段による検索結果、関連語抽出手段で抽出した関連語および検索を行った日付を少なくとも検索結果履歴として記憶する履歴記憶手段と、履歴記憶手段に記憶された検索結果履歴を読み出して画面表示する表示手段と、を備えたため、情報分析や新たな価値の創出に寄与することができる。

【0064】また、本発明のコンピュータ読み取り可能な記録媒体(請求項2)によれば、請求項1に記載の文書検索装置の各手段としてコンピュータを機能させるためのプログラムを記録したため、このプログラムをコンピュータに実行させることにより、情報分析や新たな価値の創出に寄与することができる文書検索装置を実現することができる。

【図面の簡単な説明】

【図1】本実施の形態の文書検索装置のシステム構成図である。

【図2】本実施の形態の文書検索装置において、検索ソフトの処理を示す概略ブロック図である。

【図3】本実施の形態の文書検索装置において、転置ファイルの生成処理を示すフローチャートである。

【図4】本実施の形態の文書検索装置において、文書の検索処理を示すフローチャートである。

【図5】本実施の形態の文書検索装置において、関連語の抽出処理を示すフローチャートである。

【図6】本実施の形態の文書検索装置において、検索結果履歴の一例を示す説明図である。

【符号の説明】

100 クライアント

11

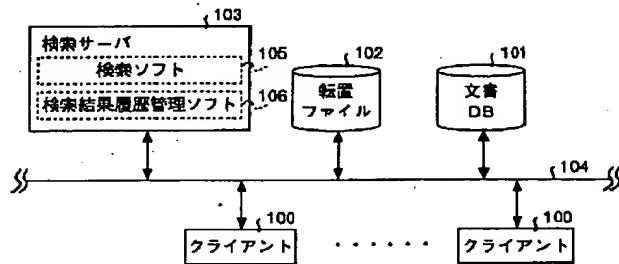
101 文書DB
 102 転置ファイル
 103 検索サーバ
 104 ネットワーク
 105 検索ソフト
 106 検索結果履歴管理ソフト
 200 自然言語処理モジュール
 201 辞書
 202 文法辞書
 204 ドキュメント・セット

12

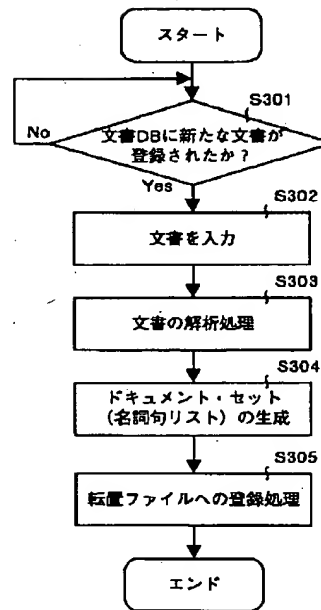
205 データベース・ビルド・コンポーネント
 206 検索条件
 207 クエリー・ビルド・コンポーネント
 208 クエリー・ドキュメント
 209 検索エンジン
 210 文書リスト
 211 関連語抽出エンジン
 212 関連語リスト
 600 検索結果履歴

10

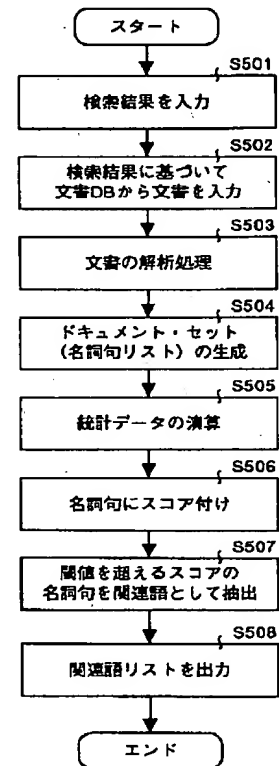
【 図1 】



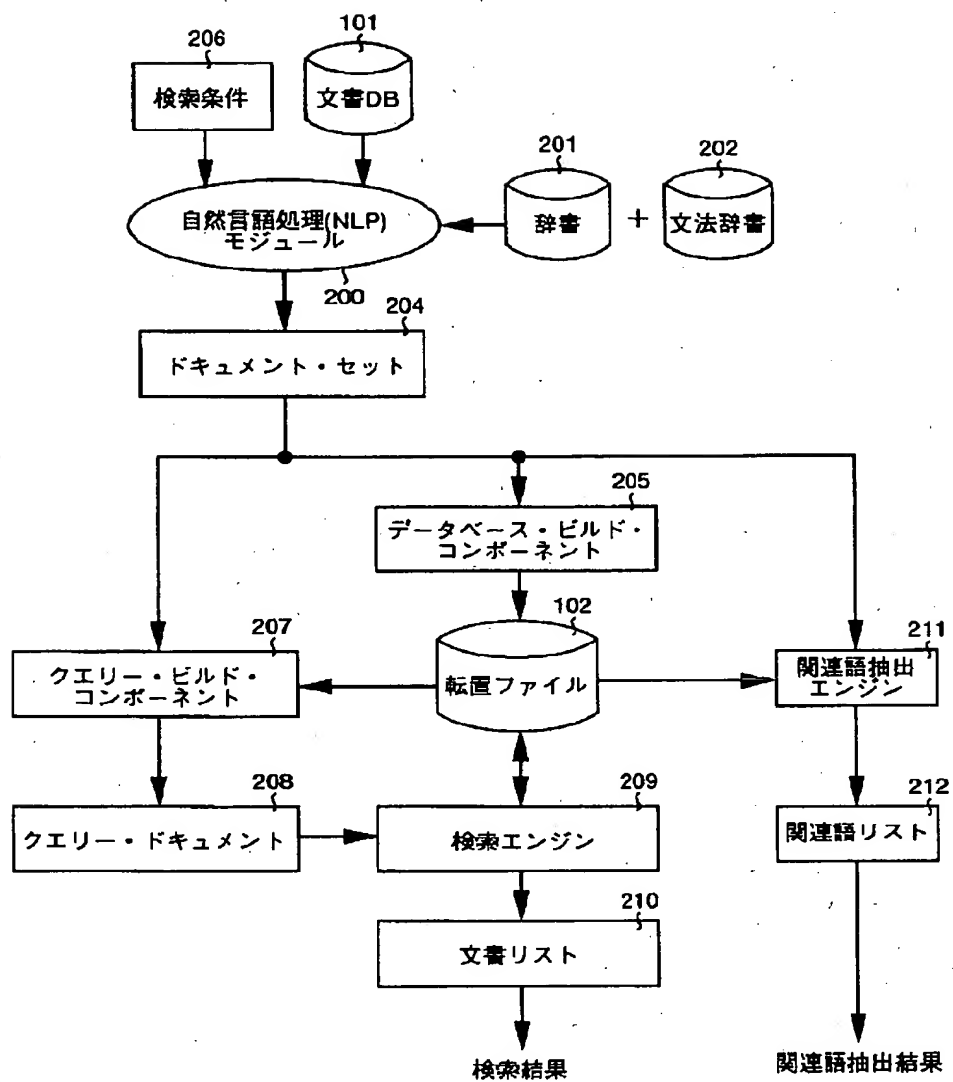
【 図3 】



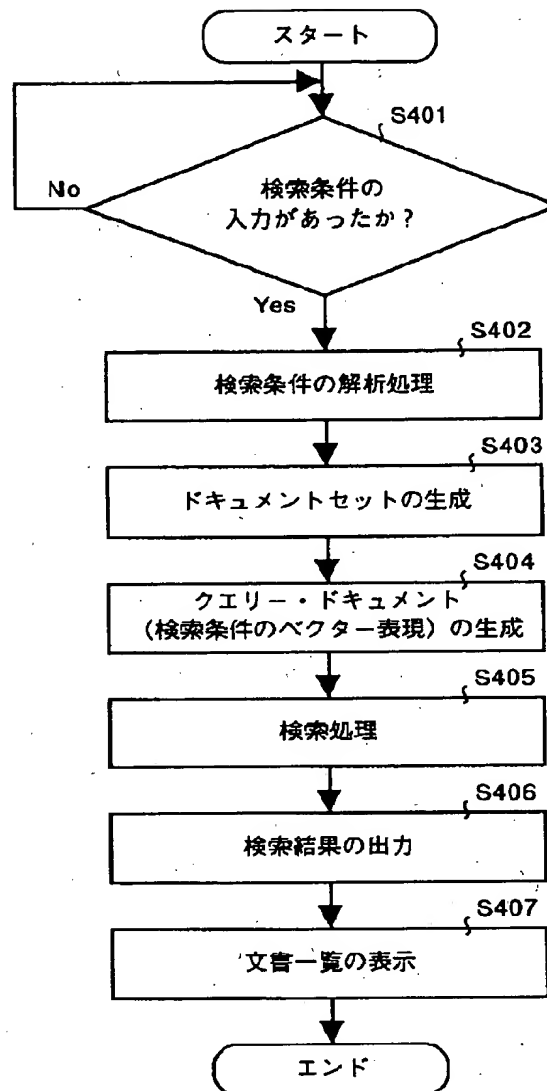
【 図5 】



【 図2 】



【 図4 】



【 図6 】

600

検索条件	検索対象	関連語	検索結果	関連文書	日付	検索者	コメント	ビュー条件	自動検索条件
米国の小麦の収獲	news96.db	米国、小麦、収獲、 小麦輸出、小麦輸入、 穀物...	komugi1.jid komugi2.jid	kazan.jid	96/6/13	山田	火山の噴火と 小麦収獲の関係	日付	半年に一度
...

フロント ページの続き

(72)発明者 荒井 好哉
 徳島市沖浜東3 丁目46番地 株式会社ジャ
 スト システム内

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.